INSTITUTO SUPERIOR TÉCNICO

# 5º Laboratório de Sistemas e Sinais

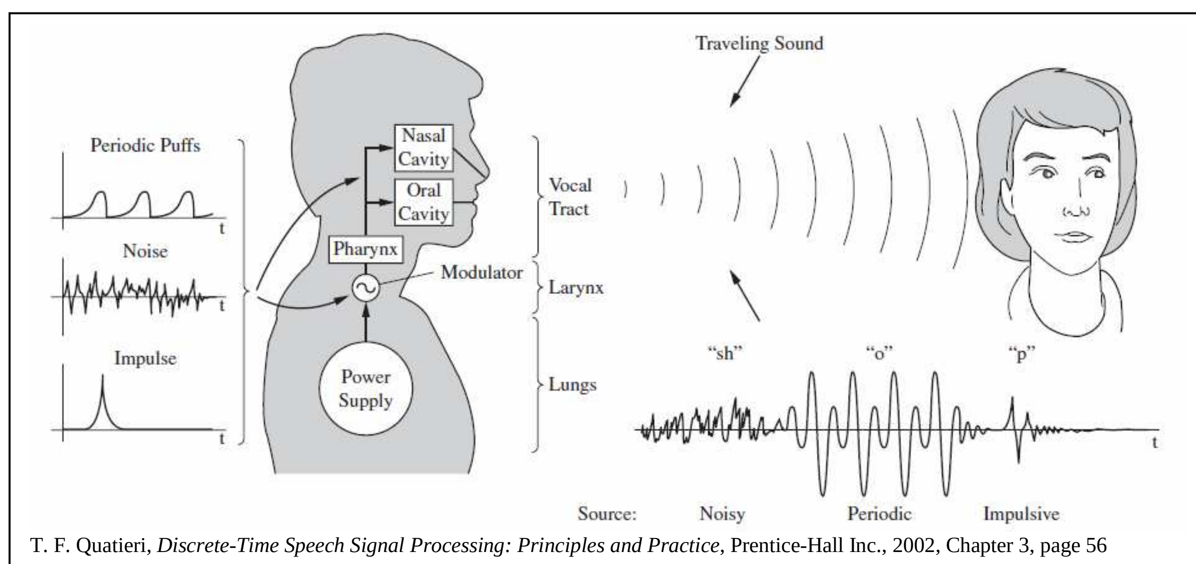
(LEIC – Alameda – 2007/08)

**Data de realização e de entrega:** semana 26 a 30 Nov de 2007.
**Local da realização:** Laboratório de Controlo, Automação e Robótica, localizado no piso 1 (cave) do Pavilhão de Mecânica III.
**Relatórios:** Os relatórios seguem a estrutura descrita na secção *Aulas de Laboratório* do site de SS no fenix. Os ficheiros resultantes devem ser comprimidos num único ficheiro, cujo nome segue a norma ***SS_5_#grupo***. A entrega do ficheiro é feita na própria aula. O laboratório está cotado em 19 valores. A qualidade do relatório está cotada em 1 valor.

---

## Exercício 1

A figura seguinte é uma representação simplificada dos mecanismos utilizados para a produção da voz humana. O sistema de *produção de voz* integra um conjunto de subsistemas fisiológicos que interagem para a produção de sinais sonoros estruturados e inteligíveis.



T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice-Hall Inc., 2002, Chapter 3, page 56

O contributo de cada subsistema fisiológico pode ser resumido como se segue:

- Os pulmões fornecem o fluxo de ar necessário à produção de um determinado som.
- A laringe, que é constituída por um conjunto de músculos, cartilagens e ligamentos, faz a modulação do som pela acção combinada de dois mecanismos: a variação da secção nas cordas vocais, o que permite controlar o caudal de ar que é fornecido ao trato vocal; a vibração do ar devida ao atrito entre as membranas das cordas vocais e o caudal de ar que cruza a superfície de contacto. Esta modulação do ar produz os sons que correspondem às vogais, como é o caso do fonema "o" na figura anterior, e outros fonemas de carácter periódico.
- O trato vocal, que integra a faringe e as cavidades nasal e bocal, "tempera" o som proveniente da laringe pela modificação selectiva de certas componentes no espectro de frequências.

A compreensão dos fenómenos fisiológicos inerentes à fala humana é importante para o desenvolvimento de formas automáticas para o processamento de voz em sistemas que interagem

com humanos. As aplicações de voz são muito diversas incluindo áreas como o desenvolvimento de interfaces homem-máquina para aplicações multimédia ou o projecto de próteses medicinais, por exemplo aparelhos auditivos ou sintetizadores de voz para traqueotomizados.

O processamento da voz por meios automáticos engloba a *Percepção/Análise* e a *Síntese*/*Produção* dos sinais sonoros e é uma área de reconhecido interesse no processamento de sinal. Para a caracterização do sinal de voz recorre-se aos fonemas, i.e. os elementos atómicos da voz, sendo estes normalmente classificados em: vogais, semivogais e consoantes. As vogais {a, e, i, o, u} são produzidas pela acção combinada de vibrações nas cordas vocais, e por isso o sinal resultante tem um carácter periódico com frequência fundamental $F_0$, com a configuração dos lábios.

Por essa razão estes são os elementos de voz mais fáceis de classificar, sendo cada vogal caracterizada por três frequências {$F_1$, $F_2$, $F_3$} denominadas por Formantes. As frequências {$F_1$, $F_2$} correspondem às frequências de ressonância das cordas vocais. A frequência $F_3$ é determinada pela abertura/forma dos lábios. As assinaturas das Formantes no espectro de frequências são assinaladas pelos três picos principais no gráfico do espectro, que correspondem às frequências de ressonância.

i) Caracterize cada uma das vogais {a, e, i , o, u} pela estimação dos seguintes parâmetros: frequência fundamental do sinal $F_0$; frequências das Formantes {$F_1$, $F_2$, $F_3$}. Utilize para o efeito os ficheiros *a.wav*, *e.wav*, *i.wav*, *o.wav* e u*.wav*, que são fornecidos no ficheiro *dados.zip*. *(3.0 valores)*

   (**Sugestão:** utilize a função `som_spectrum.m` para a análise estática e a função `som_smlnk.mdl` para visualizar a análise espectral do sinal no tempo – tenha o cuidado de usar a variável correcta no bloco `"Som"` e de escolher o tempo de simulação adequado. Ambas são fornecidas no ficheiro *dados.zip.* Esta sugestão é extensível às alíneas seguintes.)

ii) Represente num gráfico os valores de $F_1$ em função de $F_2$, i.e. um gráfico a 2 dimensões onde $F_1$ aparece no eixo das ordenadas e $F_2$ aparece no eixo das abcissas. Dê a cada ponto o rótulo correspondente à vogal respectiva e comente a disposição relativa destes pontos tendo em conta o conteúdo em frequência de cada vogal. *(1.0 valores)*

iii) Analise os valores obtidos para a formante $F_3$. O que conclui acerca da importância desta formante para a caracterização das vogais em comparação com o par {$F_1$, $F_2$}. *(1.0 valores)*

iv) Implemente uma função em Matlab para detectar as localizações das Formantes em cada uma das janelas do espectro de frequências e fazer a sua classificação. A classificação do fonema deve ser feita com o rótulo da vogal correspondente, ou com o símbolo nulo caso caso naõ se encontre uma correspondência entre o fonema e os valores estimados na alínea i). A função deve aceitar como argumento o sinal de som e produzir uma saída pertencente ao seguinte alfabeto {a, e, i, o, u, nulo}. *(5.0 valores)*

   (**Sugestão:** utilize a função `spectrogram` do matlab)

v) Aplique a função implementada na alínea iv) ao caso da transcrição automática de conteúdo sonoro para a forma escrita. Utilize como entradas os ficheiro *som_1.wav*, *som_2.wav* e *som_3.wav*. Comente os resultados obtidos. *(5.0 valores)*

*vi)* Aplique a função implementada na alínea iv) ao caso da detecção de uma sequencia de vogais particular, por exemplo 'iae', num discurso. Utilize como entradas os ficheiro *som_4.wav* e *som_5.wav*. Comente os resultados obtidos. *(4.0 valores)*

**Sugestões de leitura complementar:**

- http://www.sipl.technion.ac.il/new/Teaching/Courses/Israel_Cohen/Digital_Speech_Processing/2_Prod uction_and_Classification_of_Speech_Sounds.pdf (Atenção: à quebra do endereço!)
- http://svr-www.eng.cam.ac.uk/%7Eajr/SpeechAnalysis/index.html